Information Hiding in Images Using Steganography Techniques

Ramadhan Mstafa[1], Christian Bach[2]

Abstract- Innovation of technology and having fast Internet make information to distribute over the world easily and economically. This is made people to worry about their privacy and works. Steganography is a technique that prevents unauthorized users to have access to the important data. The steganography and digital watermarking provide methods that users can hide and mix their information within other information that make them difficult to recognize by attackers. In this paper, we review some techniques of steganography and digital watermarking in both spatial and frequency domains. Also we explain types of host documents and we focused on types of images.

INTRODUCTION

The Internet is an innovation technology that has become one of the most important events in modern world history [1]. It contains huge amounts of information in different fields. People who have a computer can get information that related to their fields without any difficulty [8]. As a result, each user who has an internet connection can read up-to-date news on the Internet, watch movies, get books, contact universities, purchase goods, etc [11]. Digital multimedia is data that can distribute easily over the Internet, making many copies of this data, breaking the intellectual property (IP) rights by authorized users more than ever. Thus, owners of those data are thinking for new technologies that promise to protect their rights [3; 5; 7].

Due to the rapid innovation of software programming on the Internet in the past two decades, there has been increasing interest in ways of hiding information in other information [12]. Many techniques are available to prevent unauthorized users from copying information without owner permission [30; 34]. Two of these techniques are cryptography and steganography [21; 22]. Cryptography is a rule or protocol between transmitter and receiver using some encryption keys to understand each other. Those encryption keys can be private (the user can make one) or public. Unauthorized users can see the coded information without understanding or being able to read it [17; 18; 19; 21]. The second method is steganography, which is embedded information which does not appear to users [23].

SCIENTIFIC RESEARCH METHOD

Researchers can make conclusions after studying the resolvable problem and having their results. This process will be done by applying scientific methods which consists number of steps [33]. Using those scientific steps leads to make the results more reasonable and proven. All scientific community members will have tested the hypothesis. Those steps are hypotheses, predictions, and test of predictions which are the life cycle of scientific

1 Department of Computer Science and Engineering, School of Engineering, University of Bridgeport, 126 Park Avenue, Bridgeport, CT 06604, rmstafa@my.bridgeport.edu

2 Department of Technology Management, Department of Biomedical Engineering, School of Engineering, University of Bridgeport, 221 University Avenue, room 153 Technology Building, Bridgeport, CT 06604, cbach@bridgeport.edu

research method. The scientific research steps like basis that someone can build its research on it. It's difficult to guide any research without having strong basis [9].

STEGANOGRAPHY

Steganography comes from the combination of the Greek words Stegano means sealed and Graphy referring to writing which means secret writing. Steganography is a very old art of embedding personal information into other data by using some rules and techniques [23]. As a result, unauthorized users are not able to see and recognize the embedded information. Steganography is managing a secret path for sending information invisibly. Figure1 shows two general directions of steganography: protection against detection and protection against removal [24].

Protection against detection uses some ways to embed information invisibly that does not degrade the quality of the original data. Protection against removal supposes that the method should be able to resist to common digital signal processing and noises. Removing the hidden data will definitely reduce the object's quality and its performance will not be functional [26; 27].

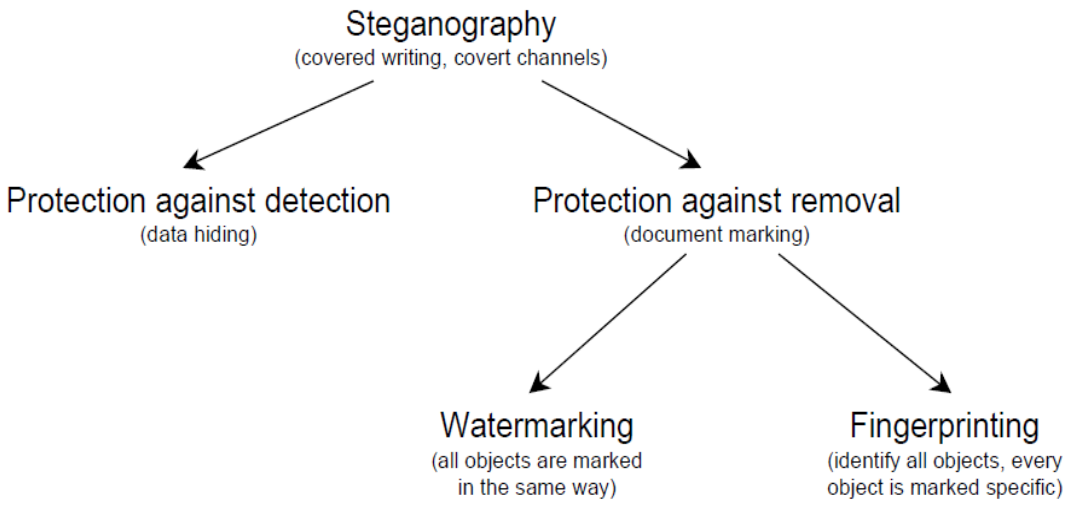

Figure1. Direction within steganography [31] p. 2.

DIGITAL WATERMARKING

Digital watermarking is one of the most widely used applications for steganography technique. Watermarking hides information in a digital signal. It is often unchangeable with text, audio, pictures, or video, for use as copy protection. So the secret information would mixes with the original signal. For example, if the signal is copied, then the information is also carried in the copy [4].

Digital Watermarking Applications

There are many applications in which digital watermarking can be used. Digital watermarking applications are important which is the requirements of digital are based on them. Applications are such as copy prevention or control, fingerprinting, broadcast monitoring, identification card security, Fraud and tamper detection, data authentication, ownership assertion, and medical applications. Some of those are listed below [42]:

- Copy Protection: Called as copy control, to prevent unauthorized users to have making illegal duplicate copies of copyrighted content.

- Owner Identification: It is the same as copyright protection, to set up ownership of the content.
- Content Authentication: To discover the modifications of the content, as a mark of incorrect authentication.
- Fingerprinting: Sometimes described as tracking of transaction or tracking of traitor, to find back distribution of the content and illegal duplication.
- Broadcast Monitoring: Used in entertainment and for advertisements and in many industries. The purpose is that to monitor data that have to be broadcasted as contracted and by the authorized user.
- Medical Applications: It is very important in medical fields also called as invertible watermarking; it contents both confidentiality and authentication in a reversible manner without having to affect the medical image in any side.

Types of Digital Watermarking

According to visibility, there are two types of digital watermarking: visible and invisible. In a visible watermarking, data is visible in the image or video. Usually the information is a text message or a company logo which recognizes the owner of the media. Most television channels have logos that indicate that the information on the specific channel is protected. Nobody is allowed to use this data without permission from the channel that owns the data. The logo means a visible watermark that can be added [40].

An invisible watermarking is information added to a digital multimedia object such as a text, audio, image, or video. An object that contains an "invisible watermark" should look like the original object. One of the most important applications of an "invisible watermark" is copyright protection. It is useful as a way of recognizing the author, creator, owner, and authorized client of a document or information [35].

Explanation of Images

As a matter of fact, a computer manipulates images as a group of picture elements called pixels. Each pixel represents a stream of binary numbers that express the pixel's intensity or color [40].

According to the color, images can be categorized into two kinds of images. One is a grayscale image, in which each pixel has 8 bits (1 byte) and the second is color image, in which each pixel has 24 bits (3 bytes). The 8-bit image has 256 different gray palettes (2^8=256). This type of image will be displayed as a black-and-white picture (0 refers to black and 255 is white). A 24-bit image consists of three fundamental colors: "red, green, and blue" (RGB); each pixel is represented by three bytes. Each byte refers to the intensity of the three main colors RGB, respectively. This type of image has good quality, and the number of palettes is more than 16 million (2^{24}) different color [31].

According to extensions, images are divided into many types such as JPEG (Joint Photographic Experts), BMP (Bitmap), PNG (Portable Network Graphics), GIF (Graphics Interchange Format), TIFF (Tagged Image File Format), and etc. Most of these extensions use RGB format to show intensity of pixel color. The web page programming such as Hypertext Markup Language (HTML) uses RGB, where each two hexadecimal digits represent one primary color. This means each pixel has six hexadecimal digits. For example, the color yellow can be created by a full amount of red color (decimal 255, hex FF); the full amount of green, the pixel's value will be "#FFFF00" in the hexadecimal system number [28; 31].

Images are of different sizes, which depend totally on the number of pixels and also on the number of bits in each pixel. The size of an 8-bit gray image consists of resolution 320 by 240 pixels which is equal to 75 Kilobytes (320*240 bytes), while the size of an image with a full color (24-bit RGB) is going to be 225 Kilobytes [4; 14; 41].

It is necessary to reduce image file sizes when transmitting via the internet. For this purpose many compression methods were developed over recent years. The two most popular types of compression are lossy and lossless compression, which are widely used in image processing. Compression processes are especially useful in BMP, GIF, and JPEG file image types [6; 14].

Lossy compression scheme uses by JPEG images this technique try to expand the file near to the size of original file [43]. On the other hand, lossless compression is a scheme that uses to rebuild the original image by applying some software. GIF and 8-bit BMP are two types of images which use for this scheme [25; 37].

Watermarking Techniques

Spatial Domain Watermarking

There are many algorithms using original data, such as video, image, audio, and text, to hide specific information like logos or personal signatures in a spatial domain. In other words, if the original data is an image, processing would be into the pixel values without changing the data into another domain. The widest and simplest method in spatial domain is Least Significant Bit (LSB), which is replacing the first bit in each pixel by information that intends to hide [13].

Least Significant Bit Watermarking

LSB is the one of the oldest and simplest algorithms that allows users to hide their information using spatial domain [16; 38]. The human eye cannot recognize the difference that occurs in the two first bits in each pixel. In other words, the change in the least significant bit does not affect the image's quality. 24-bit images have three LSB because each RGB channel has its own LSB [15; 29]. This provides users with more storage capacity to embed the information that is necessary to hide. For example, two pixels of an RGB image color will provide six bits for watermarking. To encode a message (100111) in RGB image needs two LSB pixels [13; 31].

RGB Pixel 1 (R: 0001010<u>1</u> G: 1100110<u>0</u> B: 1110110<u>0</u>)

RGB Pixel2 (R: 1101111<u>1</u> G: 0001000<u>1</u> B: 1100100<u>1</u>)

To hide the same message (100111) in a gray-scale image six LSB pixels are needed.

Pixel1: 1001010<u>1</u> Pixel2: 0000110<u>0</u> Pixel3: 1100100<u>0</u>

Pixel4: 1001111<u>1</u> Pixel5: 0001000<u>1</u> Pixel6: 1100101<u>1</u>

Frequency Domain Watermarking

This is also called transform domain, because the original data changes from spatial to frequency domain. The most common frequency methods are Discrete Fourier Transformation (DFT), Discrete Wavelet Transformation (DWT), and Discrete Cosine Transformation (DCT). For example, an 8-bit image with a 256 by 256 resolution can be transformed into frequency watermarking using DWT. The result of this processing would be four small images, each of them with a 128 by 128 resolution. Moreover, four images will have different frequency ranges from low to high because each of them has different coefficients for others. The main advantage of using frequency domain watermarking is that it is robust for many kinds of signal manipulations when sending data via the Internet. Also, it resists of many noises that attack embedded information [2; 13; 31].

Discrete Wavelet Transform

It is a tool to transform the signal or data from one domain which is a spatial to another domain which is a frequency. In the frequency domain the signal splits into the two half one of them is high frequency and another is low frequency. Then each of them is going to divide again into high and low frequency that four different parts of signal [10]. Four parts or sub bands of decomposed signal are LL, LH, HL and HH frequencies which are low-low, low-high, high-low and high-high frequencies [10; 29]. Low frequency is the same of original signal and other parts are more details of signal they are not exact data as original one, so we can change or remove depends on the technique that we using. The reconstruction process is the opposite of decomposition process that means the four bands of divided data have to be mixed again to recover the original data. Sometimes we do more than one level of decomposition depends on the algorithm that we use. Low-low frequency band will be used in case we do second decomposition. In case of reconstruction the last level of decomposition will used first which is an exact opposite direction [42].

Discrete Cosine Transform

In this tool the data will divide into some blocks often 8 by 8 or 16 by 16 blocks. Then, applying a discrete cosine transform on each block will convert the signal into high, middle and low frequencies. Low frequency is very close to original data while the middle and high frequencies are more details of the data. We can use the details frequencies as a host data to hide some important secret on it or we can remove those details frequencies to reduce the size of the signal. The reconstruction process is rebuilding the signal in opposite way it means combining all frequencies high, middle and low into single signal [32].

Embedding and Detection Processes

In embedding process the secure data which called the logo will be embedded into the host data sometimes call cover data and send to the destination. User can use many secret keys; in general we can divide into two types. First, symmetric key which both sender and receiver have the same key for encryption and decryption data. Second, asymmetric key both transmitter and receiver use different types of keys. Watermarked data is the data that has to be sent to destination which consists of mixing logo, cover and key data which seems to anyone that is one piece of data [39]. Figure2 shows embedding process.

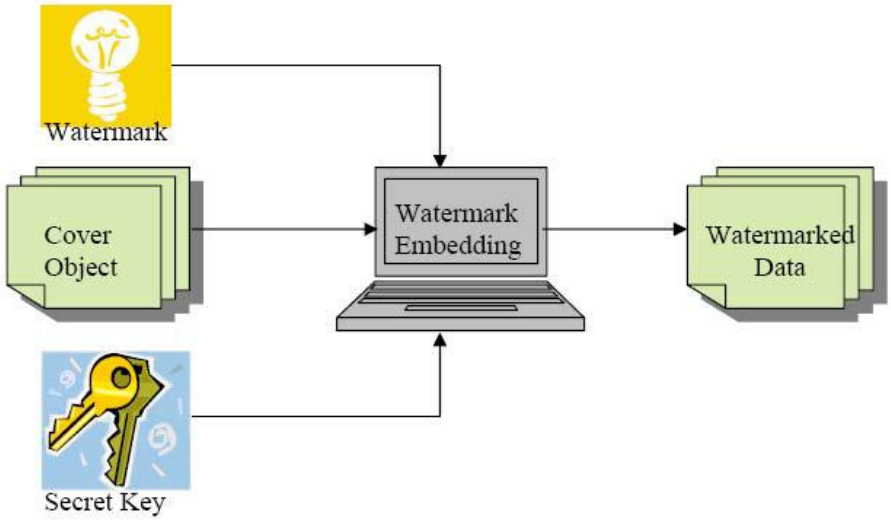

Figure2. Watermark Embedding Process [42] p. 2.

In detection process, when the watermarked data reaches to the destination as one piece of data which in reality it is a group of mixed data. The logo data will be extracted from the mixed data by using one type of key. Splitting of those three signals needs to use one of techniques in both spatial and frequency domains. The extraction process depends on the type of the algorithm that used and the quality of recovered signals is different from using one algorithm to others. Also the number of decomposition levels that used in embedding process affects directly to the quality of the data that have been sent it by user which is using the same number of reconstructions levels [36]. Figure3 shows detection process.

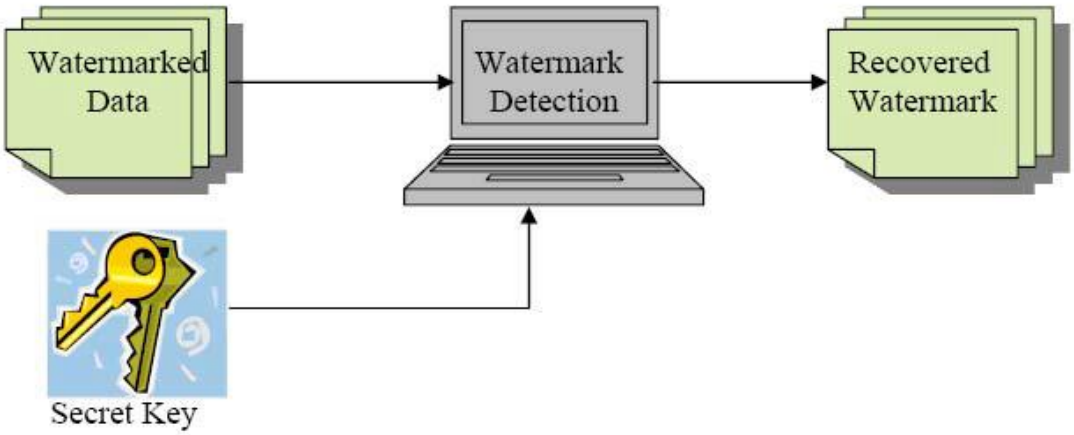

Figure3. Watermark Detection Process [42] p. 2.

CONCLUSION

Although there are many advantages of the internet, it has also opened a new way for invasion of our privacy and intellectual property by hackers and unauthorized users. Many techniques have been invented since these problems appeared. One useful technique to protect information via the internet is steganography. Digital watermarking is one of the popular applications for steganography. Users can hide important information within an image by using an invisible watermark when they transmit data. Moreover, a visible watermark can be used in many applications such as author, creator, and document. Images have some unimportant regions the human visual system cannot recognize by replacing these regions with other information. A user can change the least significant bit in each pixel with his/her own information without the quality of an image being decreased. Also, this alteration does not affect the intensity of the color.

REFERENCES

[1] Afrakhteh, M., & Ibrahim, S. (2010, 25-27 June 2010). *Adaptive steganography scheme using more surrounding pixels.* Paper presented at the Computer Design and Applications (ICCDA), 2010 International Conference on.

[2] Ahmed, A. M., & Day, D. D. (2004). Applications of the naturalness preserving transform to image watermarking and data hiding. *Digital Signal Processing, 14*(6), 531-549. doi: 10.1016/j.dsp.2004.08.002

[3] Al-Hunaity, M. F., El-Emary, I. M., & Najim, S. A. (2007). Colored digital image watermarking using the wavelet technique. [Article]. *American Journal of Applied Sciences, 4*(9), 658+.

[4] Al-Otum, H. M., & Samara, N. A. (2010). A robust blind color image watermarking based on wavelet-tree bit host difference selection. *Signal Processing, 90*(8), 2498-2512. doi: 10.1016/j.sigpro.2010.02.017

[5] Alturki, F., & Mersereau, R. (2001, Apr 2001). *A novel approach for increasing security and data embedding capacity in images for data hiding applications.* Paper presented at the Information Technology: Coding and Computing, 2001. Proceedings. International Conference on.

[6] Amat, P., Puech, W., Druon, S., & Pedeboy, J. P. (2010). Lossless 3D steganography based on MST and connectivity modification. *Signal Processing: Image Communication, 25*(6), 400-412. doi: 10.1016/j.image.2010.05.002

[7] Awwad, W. F., Mansour, R. F., & Mohammed, A. A. (2012). A robust method to detect hidden data from digital images. [Report]. *Journal of Information Security, 3*(2), 91+.

[8] Babu, K. S., Raja, K. B., Kiran, K. K., Manjula Devi, T. H., Venugopal, K. R., & Patnaik, L. M. (2008, 19-21 Nov. 2008). *Authentication of secret information in image Steganography.* Paper presented at the TENCON 2008 - 2008 IEEE Region 10 Conference.

[9] Bailey, K., & Francis, M. (2008). Managing information flows for improved value chain performance. *International Journal of Production Economics, 111*, 2-12.

[10] Chandra, M., & Pandey, S. (2010, 1-3 Aug. 2010). *A DWT domain visible watermarking techniques for digital images.* Paper presented at the Electronics and Information Engineering (ICEIE), 2010 International Conference On.
[11] Chang, C.-C., Chen, W.-J., & Le, T. H. N. (2010). High payload steganography mechanism using hybrid edge detector. [Report]. *Expert Systems With Applications, 37*(4), 3292+.
[12] Chang, C.-C., Chuang, J.-C., & Lin, P.-Y. (2010). A grayscale image steganography based upon discrete cosine transformation. [Technical report]. *Journal of Digital Information Management, 8*(2), 88+.
[13] Cheddad, A., Condell, J., Curran, K., & Mc Kevitt, P. (2010). Digital image steganography: Survey and analysis of current methods. *Signal Processing, 90*(3), 727-752. doi: 10.1016/j.sigpro.2009.08.010
[14] Chen, W.-Y. (2007). Color image steganography scheme using set partitioning in hierarchical trees coding, digital Fourier transform and adaptive phase modulation. *Applied Mathematics and Computation, 185*(1), 432-448. doi: 10.1016/j.amc.2006.07.041
[15] Chin-Chen, C., & Hsien-Wen, T. (2009, 4-6 June 2009). *Data Hiding in Images by Hybrid LSB Substitution.* Paper presented at the Multimedia and Ubiquitous Engineering, 2009. MUE '09. Third International Conference on.
[16] Ching-Sheng, H., & Shu-Fen, T. (2010, 26-28 Feb. 2010). *Finding Optimal LSB Substitution Using Ant Colony Optimization Algorithm.* Paper presented at the Communication Software and Networks, 2010. ICCSN '10. Second International Conference on.
[17] El-Emam, N. N. (2007). Hiding a large amount of data with high security using steganography algorithm. [Article]. *Journal of Computer Science, 3*(4), 223+.
[18] Farshchi, S. M. R., & Toosizadeh, S. (2011). High secure communication using chaotic double compression steganography technique. [Report]. *International Journal of Research and Reviews in Computer Science*, 527+.
[19] Hedieh, S., & Jamzad, M. (2008, 8-11 July 2008). *Cover Selection Steganography Method Based on Similarity of Image Blocks.* Paper presented at the Computer and Information Technology Workshops, 2008. CIT Workshops 2008. IEEE 8th International Conference on.
[20] Hung-Min, S., King-Hang, W., Chih-Cheng, L., & Yih-Sien, K. (2007, Oct. 30 2007-Nov. 2 2007). *A LSB substitution compatible steganography.* Paper presented at the TENCON 2007 - 2007 IEEE Region 10 Conference.
[21] Husainy, M. A. F. A. (2009). Image steganography by mapping pixels to letters. [Report]. *Journal of Computer Science, 5*(1), 33+.
[22] Ibrahim, B., Jabri, R., & Zoubi, H. A. (2009). Information hiding: a generic approach. [Technical report]. *Journal of Computer Science, 5*(12), 933+.
[23] Jin-Suk, K., Yonghee, Y., & Mee Young, S. (2007, 7-9 Nov. 2007). *Steganography using block-based adaptive threshold.* Paper presented at the Computer and information sciences, 2007. iscis 2007. 22nd international symposium on.
[24] Li, B., Biswas, S., & Blasch, E. P. (2007, 9-12 July 2007). *An estimation approach to extract multimedia information in distributed steganographic images.* Paper presented at the Information Fusion, 2007 10th International Conference on.
[25] Li, L.-d., Guo, B.-l., & Guo, L. (2008). Rotation, scaling and translation invariant image watermarking using feature points. *The Journal of China Universities of Posts and Telecommunications, 15*(2), 82-87. doi: 10.1016/s1005-8885(08)60089-8
[26] Martin, A., Sapiro, G., & Seroussi, G. (2005). Is image steganography natural? *Image Processing, IEEE Transactions on, 14*(12), 2040-2050. doi: 10.1109/tip.2005.859370
[27] Marvel, L. M., Retter, C. T., & Boncelet, C. G., Jr. (1998, 4-7 Oct 1998). *Hiding information in images.* Paper presented at the Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on.
[28] McBride, B. T., Peterson, G. L., & Gustafson, S. C. (2005). A new blind method for detecting novel steganography. *Digital Investigation, 2*(1), 50-70. doi: 10.1016/j.diin.2005.01.003
[29] Min-Jen, T., & Jung, L. (2011, 6-9 Nov. 2011). *The quality evaluation of image recovery attack for visible watermarking algorithms.* Paper presented at the Visual Communications and Image Processing (VCIP), 2011 IEEE.
[30] Neeta, D., Snehal, K., & Jacobs, D. (2007, 6-6 Dec. 2006). *Implementation of LSB Steganography and Its Evaluation for Various Bits.* Paper presented at the Digital Information Management, 2006 1st International Conference on.

[31] Popa, R. (1998). An analysis of steganographic techniques. *The Politehnica University of Timisoara, Faculty of Automatics and Computers, Department of Computer Science and Software Engineering*.
[32] Rongsheng, X., Keshuo, W., & Shunzhi, Z. (2007, 16-18 April 2007). *An Improved Semi-fragile Digital Watermarking Scheme for Image Authentication.* Paper presented at the Anti-counterfeiting, Security, Identification, 2007 IEEE International Workshop on.
[33] Sarg, S. (2008). Gravito-inertial Propulsion Effect Predicted by the BSM. *Supergravitation Unified Theory*.
[34] Shaohui, L., Hongxun, Y., & Wen, G. (2004, 5-7 April 2004). *Steganalysis of data hiding techniques in wavelet domain*. Paper presented at the Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on.
[35] Su, J. K., Hartung, F., & Girod, B. (1998). Digital watermarking of text, image, and video documents. *Computers & Graphics, 22*(6), 687-695. doi: 10.1016/s0097-8493(98)00089-2
[36] Subbarayan, S., & Karthick Ramanathan, S. (2009, 28-30 Dec. 2009). *Effective Watermarking of Digital Audio and Image Using Matlab Technique.* Paper presented at the Machine Vision, 2009. ICMV '09. Second International Conference on.
[37] Suhail, M. A., Obaidat, M. S., Ipson, S. S., & Sadoun, B. (2003). A comparative study of digital watermarking in JPEG and JPEG 2000 environments. *Information Sciences, 151*(0), 93-105. doi: 10.1016/s0020-0255(02)00291-8
[38] Suk-Ling, L., Kai-Chi, L., Cheng, L. M., & Chi-Kwong, C. (2006, Aug. 30 2006-Sept. 1 2006). *Data Hiding in Images by Adaptive LSB Substitution Based on the Pixel-Value Differencing.* Paper presented at the Innovative Computing, Information and Control, 2006. ICICIC '06. First International Conference on.
[39] Tripathi, S., Jain, R. C., & Gayatri, V. (2006, 0-0 0). *Novel DCT and DWT based Watermarking Techniques for Digital Images.* Paper presented at the Pattern Recognition, 2006. ICPR 2006. 18th International Conference on.
[40] Tsai, P., Hu, Y.-C., & Chang, C.-C. (2004). A color image watermarking scheme based on color quantization. *Signal Processing, 84*(1), 95-106. doi: 10.1016/j.sigpro.2003.07.012
[41] Yu, Y.-H., Chang, C.-C., & Lin, I.-C. (2007). A new steganographic method for color and grayscale image hiding. *Computer Vision and Image Understanding, 107*(3), 183-194. doi: 10.1016/j.cviu.2006.11.002
[42] Yusof, Y., & Khalifa, O. O. (2007, 14-17 May 2007). *Digital watermarking for digital images using wavelet transform.* Paper presented at the Telecommunications and Malaysia International Conference on Communications, 2007. ICT-MICC 2007. IEEE International Conference on.
[43] Zhang, X., Wang, S., Qian, Z., & Feng, G. (2010). Reversible fragile watermarking for locating tampered blocks in JPEG images. *Signal Processing, 90*(12), 3026-3036. doi: 10.1016/j.sigpro.2010.04.027

Ramadhan Mstafa
Ramadhan Mstafa, originally from Dohuk, Kurdistan Region, Iraq, is pursuing his Doctorate in Computer Science and Engineering at the University of Bridgeport in Bridgeport, Connecticut. He received his Bachelor's degree in Computer Science from the University of Salahaddin and his Master's degree in Computer Science from University of Duhok. His research interests include image processing, mobile communication, and steganography.

Christian Bach

Christian Bach received his MBA and PhD in Information Science from University at Albany SUNY in Albany, New York. Some of Dr. Bach's research interests include Intracellular Immunization, induced Pluripotent Stem (iPS) cells, Artificial Transcription Factors, Target Detection Assay, Microarrays, Bioreactors, Protein Folding (micro -level), Target Binding Site Computation, micro Database Systems, and Knowledge Cubes. He is the author of multiple journal articles including "Tower Computing: Utilization of Cloud Computing in science-based Knet environments," "Employing the Intellectual Bandwidth Model for Measuring Value Creation in Collaborative Environments," and "Scientific and Philosophical Aspects of Information and the Relationships among Data, Information, and Knowledge."